

38



Europäisches Patentamt

(19) European Patent Office

Office européen des brevets

(11) Publication number:

0 079 739

A2

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(51) Int. Cl<sup>3</sup>: C 12 N 15/00

(22) Date of filing: 08.11.82

C 12 N 1/00, C 12 P 21/02  
C 07 H 21/04, C 07 C 103/52  
//C12R1/19, C12R1/865

(30) Priority: 12.11.81 US 320632

(71) Applicant: THE UPJOHN COMPANY  
301 Henrietta Street  
Kalamazoo, Michigan 49001(US)

(43) Date of publication of application:  
25.05.83 Bulletin 83/21

(72) Inventor: Dugaiczky, Achilles  
c/o The Upjohn Company 301 Henrietta Street  
Kalamazoo Michigan 49001(US)

(84) Designated Contracting States:  
BE CH DE FR GB IT LI NL SE

(74) Representative: Perry, Robert Edward et al,  
GILL JENNINGS & EVERY 53-64 Chancery Lane  
London WC2A 1HN(GB)

(54) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(57) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

EP 0 079 739 A2

BEST AVAILABLE COPY

0079739

UPJOHN

1

GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION  
AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly  $\alpha$ -fetoprotein, but the synthesis decreases drastically after birth. Recently, Law et al determined the complete sequence of mouse  $\alpha$ -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been reached from studies on the  $\alpha$ -fetoprotein genes of the rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum mRNA has been determined from recombinant cDNA clones and from a primer-extended cDNA synthesis on the mRNA template. The sequence comprises 2,078 nucleotides, starting upstream of a potential ribosome binding site in the 5'-untranslated region. It contains all the translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the triple-domain structure of the serum albumin molecule.

DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched 5 albumin cDNA probe, and the recombinant plasmid pH A36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pH A206. The latter was obtained in a second transformation experiment after initiating 10 the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pH A36. The two plasmids, pH A36 and pH A206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending 15 into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by 20 labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pH A36, pH A206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleotides, of which 38 represent the 5'-untranslated region, 54 identify a 25 prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table 30 1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T T<sup>C</sup> T C T T C T G T.....albumin mRNA  
35 (3')...G A G G A A G G C G U C C m<sub>2</sub><sup>6</sup>A m<sub>2</sub><sup>6</sup>A.....18S RNA

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since pre-peptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous 5 translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a pre- 10 peptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the pro- 15 peptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence 20 located near the polyadenylation site has been identified by Renoist, et al. [Renoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the consensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the 25 human albumin mRNA (Table 1).

0079739

4083

-4-

TABLE 1

5      10      15      20      25      30      35

-1      -6      p      r      o      -1      1      -10

Met lys trp val thr phe ile ser leu leu phe leu phe ser

ser ala tyr ser arg gly val phe arg asp ala his lys ser glu val ala his arg phe lys asp leu ala glu asn phe lys

(30)      TCC GCT TAT TCC AGG CGT GTC TTT CGT CGA GAT GCA CAC AAG AGT GAG GTT CCT CAT CGG TTT AAA CAT TTG GCA GAA AAT TTC AAA (170)

20      25

ala leu val leu lle ala phe ala gln tyr leu gln gln cys pro phe olu esp his val lys leu val asn glu val thr glu phe ala

GCC TTG GTC TTG ATT GCC TTT GCT CAC TAT CTT CAG CAG TGT CCA TTT GAA GAT CAT GTA AAA TTA GTG AAT GAA GCA ACT CAA TTT GCA (260)

30      35      40      45      50

lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu

lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu oys thr val ala thr leu

(350)      AAA ACA TGT GTC GCT GAT GAG TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TGC ACA GAT GCA ACT CCA (440)

51      55      60      65      70      75      80

arg glu thr tyr gly glu met ala asp cys cys ala lys gln gln pro gly arg asn glu cys phe leu aln his lys asp aso arg pro

(440)      AAC GAA ACC TAT GGT GAA ATG CCT GAC TGC TGT GCA AAA CAA GAA CCT GGG AGA AAT GAA GAC ACA TTT TTG CAA CAC AAA GAT GAC AAC CCA (440)

90      91      95      100      105      110

asn leu pro arg leu val arg pro glu val asp val met oys thr ala phe his asp asn glu glu thr phe leu lys lys aso arg pro

(440)      AAC CTC CCC CGA TTC GTG AGA CCA GCA GAG GTT CAT GTC AGC ACT CCT TTT CAT GAC AAT GAA GAC ACA TTT TTG AAA AM TAC TTA TAT (330)

111      115      120      125      130      135      140

glu lle ala arg his pro tyr phe tyr ala pro glu leu leu phe ohe ala lys arg tyr lys ala phe thr glu cys oys qin

(440)      GAA ATT CCC AGA AGA CAT CCT TAC TAT GGC CCC GAA CTC CTT CCT TIC TTT CCT AAA AGG TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)

141      145      150      155      160      165      170

ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp olu ala lys ala ser ser ala lys qin aro leu lys oys

(440)      GCT CCT GAT AAA GCT GCC TGC CTC TCC CCA AAC CTC GAT GAA CCT CGG GAT GAA GCG AAC GCT TCG TCT GCC AAA CAG AGA CTC AAG TGT (710)

171      175      180      185      190      195      200

ala ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg asp olu ala olu phe ala olu

GCC AGT CTC CAA AAA TTT GGA GAA AGA CCT TTC AAA CCA TGG GCA GAA CCT CCT CGC CTC AGC CAG AGA TTT CCC AAA CCT GAG TTT GCA GAA (300)

210      215      220      225      230

ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg asp olu ala olu phe ala olu

0079739

4083

-5-

35           30           25           20           15           10           5  
231           240           245 246           250           253           259           260  
val ser lys leu val thr lys val his thr glu cys his gly asp leu leu glu cys ala asp arg ala asp leu  
CTT TCC AAG TTA GTG ACA GAT CTT ACC AAA GTC CAC ACC GAA TGC TCC CAT GCA GAT CTC CTT GAA TGT CCT GAT AAC GGC GAC CTT (890)  
261           265           270           278 279 280           278 279 280  
ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys cys alu lys pro leu leu glu lys ser his cys lle  
CCC GAA GTG GAA AAT GAT GAC ATG CCT GCT GAC TTG CCT TCA ATA GCT GAT TTT GTC GAA ACT AAC GAT CTT TGC GAA AAA TGT CAC TGT ATT (980)  
291           300           305           310           315           320  
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
GCC GCA AAG GAT GTC TGC TTC GCG ATG TTT TTG TAT GAA TAT GCA AGG CAT CCT GAT TAC TCT GTC GTC CTC AGA CCT CCC (11070)  
321           330           335           340           345           350  
glu ala lys asp val phe leu tyr glu met phe leu tyr ala aro arg his pro asp tyr ser val val leu leu arg leu ala  
GAG GCA AAG GAT GTC TGC TTC GCG ATG TTT TTG TAT GAA TAT GCA AGG CAT CCT GAT TAC TCT GTC GTC CTC AGA CCT CCC (1160)  
351           360 361           369 370           380  
lys thr tyr glu thr thr leu glu lys cys ala ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys pro leu  
AAG ACA TAT GAA ACC ACT CTA GAC AAG TCC TGT CCT GCT GCA GAT CCT CAT GAA TGC TAT GCA AAA GTC TTC GAT GAA TTT AAA CCT CCT (1250)  
381           390           392           400  
val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu aln leu aln tyrosine ala lys val leu leu arg  
CTG GAA GAG CCT CAG AAT TTA ATC AAA CAA AAC TGT GAG CTT GCA GAC TAC AAA TTC CAG AAC TCT GTC TTA GTC CGT (1340)  
411           420           430           437 438 440  
tyr thr lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu aln lys val aln ser lys cys cys lys hls  
TAC ACC AAG AAA GCA ACC CCC TGT GCA GAA GAC TAT CTA TCC AAC CAG CCT TCA AGA AAC CTA GCA AAA GTG GTC TGG CAT GAC AAA TGT TGT AAA CAT (1430)  
441           448           450           460 461           470  
pro glu ala lys arg met pro gys ala glu asp tyr leu ser val val leu asn gln leu cys val leu hls glu lys thr pro val ser  
CCT GAA GCA AAA AGA ATG CTC ACC AAA TGC TCC ACA GCA TCC TTG GTG AAC CCG TAT CTA TCC AAC CAG CCT GTC TGG CAT GAA ACA TAC GTC CCC GCA ACT (1520)  
471           476 477           480           490  
asp arg val thr lys cys thr glu ser leu val arg arg pro cys phe ser ala leu glu val asp glu thr tyr val pro lys  
GAC AGA GTC ACC ACC AAA TGC TCC ACA GCA TCC TTG GTG AAC CCG TAT CTA TCC AAC CAG CCT GTC GAT GAA GTC CCT GAT GAA ACA TAC GTC CCC AAA (1610)  
501           510           515           520           530  
glu phe asn ala glu thr phe his ala asp lle cys thr leu ser glu arg aln lle lys lys aln thr ala leu val  
GAG TTT AAT GCT GAA ACA TTC ACC TCC CAT GCA GAT ATA TGC ACA CTT TGT CAC AAC GAG AGA CAA ATC AAC ACT GCA CTT CTT GTC (1700)

0079739  
4083

-6-

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5    Example 1      Isolation of Messenger RNA

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and 1D Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, U.K. (1970) Nature 227, 15 680-685.

15    Example 2      Cloning Procedures

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczky, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Bolivar, F., 20 Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczky, A., Robberson, O.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., Ibid.]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [<sup>32</sup>P]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

30    As shown in Example 5, plasmids pH A36 and pH A2D6 were deposited in E. coli HB1D1 hosts. The plasmids were obtained from E. coli RR1 hosts, described in this example, and transformed into E. coli HB1D1 by standard procedures well known to those of ordinary skill in this art. The E. coli RR1 hosts were lysed and then centrifuged to 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris-HCl, pH 8.0, 1D mM CaCl<sub>2</sub>, 1D mM MgCl<sub>2</sub>). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HB101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml 5 chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of 10 L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Bethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) *J. Virol.* 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) *Biochemistry* 11, 1242-1250] gels.

Example 4 DNA Sequencing

DNA fragments were dephosphorylated with bacterial alkaline phosphatase (Worthington) and labeled at the 5'-ends with poly-nucleotide kinase (Boehringer-Mannheim) and  $\gamma$ [<sup>32</sup>P]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) *Methods Enzym.* 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) *FEBS Letters* 87, 107-110].

Example 5 Recombinant Plasmids pH A36 and pH A206

35 As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pH A36 contained the largest insert of an albumin cDNA sequence. Both plasmids pH A36 and pH A206 have been deposited in a viable *E. coli* host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

5 HB101(pHA206) - NRRL B-12550

E. coli HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

10 NRRL B-12550 and NRRL B-12551 are available to the public. ~~upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.~~

15 E. coli RR1 and E. coli HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

20 YEpl6 is a well known and widely available yeast episomal plasmid. It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEpl6) - NRRL B-12093.

Example 6 Assembly of the Serum Albumin Gene

25 Assembling the pieces together is a straightforward task of restriction enzymology. There is only one Mspl site in the overlapping DNA sequence of the two cDNA clones. Two enzymatic steps of (i) Mspl digestion of the two DNAs, followed by (ii) the use of ligase, an enzyme that seals DNA fragments, will give the desired product. Although two other undesired DNA species will also be obtained in the 30 course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

35 (a) Escherichia coli

(b) Saccharomyces cerevisiae

(a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the 5 existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one 10 of the yeast plasmid vectors, e.g., YEp6, at the EcoR1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the beginning 20 of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466.

Example 8 Screening of Clones Producing Albumin

35 Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an in situ lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies 5 producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. 10 Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

15

20

25

30

35

CLAIMS

1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.

5

2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.

3. E. coli HB101 (pHA36) having the deposit accession number  
10 NRRL B-12551.

4. E. coli HB101 (pHA206) having the deposit accession number  
NRRL B-12550.

15 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

30

35

0079739

4083

-13-

0079739

4083

-14-

0079739

4083

-15-

0079739

-16-

4083

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

1  
 asp ala his lys ser glu val ala his arg phe lys asp leu ala glu asn phe lys  
 GAT GCA CAC AAG AGT CAC GTT CCT CAT CGG TTT AAA GAT TTG GCA GAA GAA AAT TTC AAA (170)  
  
 21  
 ala leu val lle ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala  
 GCC TTC GTC ATT GCC TTT GCT CAC TAT CCT CAG CAC TGT CCA TTT GAA CAT CAT GTA AAA TTA GTC AAT GAA GCA ACT CAA TTT GCA (260)  
  
 51      53      60      62      66  
 lys thr cys val ala asp glu aer ala glu asn cys asp lys ser leu his thr leu phe gln asp lys leu cys thr val ala thr leu  
 AAA ACA TGT CTT GCT GAT CAC TCA GCT AAC TCA CTT CAT ACC CTT CCT TTT CGA GAC AAA TTA TGC ACA GTC AAC ACT CTT (350)  
  
 61  
 arg glu thr tyr gln glu met ala asp cys cys ala lys gln glu pro gln arg asn glu cys phe leu gln his lys asp asn pro  
 CCT GAA ACC TAT GCT GAA ATG CCT GAC TGC TGC TGT GCA AAA CAA CAA CCT GGC AGA AAT GAA TGC TTC TTG CAA CAC AAA GAT GAC AAC CCA (440)  
  
 111  
 asn leu pro arg leu val arg pro glu val met cys thr ala phe his asp asn glu glu thr phe leu lys lys tyr leu try  
 AAC CTC CCC CCA TTC GTC AGC CCA GAC CTT GAT GTC ATG TGC ACT CCT TTT CAT GAC AAC TAT GAA GAC ACA TTT TGC AAA AAC TAC TTA TAT (530)  
  
 141  
 glu lle ala arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala phe thr glu cys cys aln  
 GAA ATT CCC AGA AGA CAT CCT TAC TTT TAT GGC CCC GAA CTC CTT CTC GAA CTT CGC GAT GAA CCT AAA AGG TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)  
  
 171  
 ala ala asp lys ala ala cys leu leu pro lys leu asp glu leu arg asp glu gln gln lys ala lys ala ser ala lys ala lys cys  
 CCT CCT GAT AAA CCT GCC TGC CTC GTC CTC GAT GAA CCT CGC GAT GAA CCT CCC GCA (710)  
  
 201  
 ala ser leu gln lys phe gly glu arg ala pha lys ala trp ala val ala arg leu ser gln arg phe ala glu glu  
 GCC AGT CTC CAA AAA TTT GCA GAA AGA GCT TTC AAA GCA TGC GCA (800)

100

110

120

130

140

150

160

170

180

190

200

210

220

230

0079739  
4083

-17-

231	240	245	246	250	253	255	260
val ser lys leu val thr asp leu thr lys val his thr pro cys cys his gly asp leu leu glu cys ala asp asp arg ala asp leu							
GTT TCC AAG TTA GTC ACA CAT CTT ACC AAA GTC CAC ACC GAA TCC TGC CAT CCT GAA TGT GCA TCT GAA TGT GCT GAT GAC AGG GCG GAC CTT (890)							
261	265	270		278	279	280	289 290
ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys cys glu lys pro leu leu glu lys ser his cys lle							
GCC AAC TAT ATC TGT GAA AAT CAA CAT TCG ATC TCC AGT AAA CTC AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TGT CAC TGC ATT (980)							
291		300		310	316		320
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala							
GCC GAA GTG GAA AAT GAT GAG ATG CCT CCT GAC TTC CCT GCT GAT TTT GCT GAA AGT AAG GAT GTC TGC AAA AAC TAT CCT (1070)							
321		330		340		350	
glu ala lys asp val phe leu gly met phe leu tyr ala arg arg his pro asp tyr ser val val leu leu ala asp							
GAG GCA AAG GAT GTC TTC TGC GCC ATG TTT TGC TAT GAA TAT GCA AGA ACC CAT CCT GAT TAC TCT GTC GTC CTC CTC AGA CCT CCC (1160)							
351	360	361		369	370	380	
lys thr tyr glu thr leu glu lys cys cys ala ala ala asp pro his glu cys tyr ala lys val ala asp glu phe lys pro leu							
AAG ACA TAT GAA ACC ACT CTA CAG AAC TGC TGT GGC CCT GCA GAT CCT CAT GAA TGC TAT GGC TAT GCA AAA GTG TTC GAT GAA TTT AAA CCT CCT (1250)							
381		390	392		400		410
val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu glu tyro lys phe gln asn ala leu leu val arg							
GTC GAA GAG CCT CAG AAC TAT TTA ATC AAA CAA AAC TAT GAG CTT TTT GAC CAG TAC AAA TAC AAC CTC TTA CCT CGT TTA CCT CGT (1360)							
411		420		430		437 438	440
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu gln lys val gln ser lys cys cys lys his							
TAC ACC AAC AAA GTA CCC CAA GTG TCA ACT CCA ACT CTT GTC GAG GTC GTC TCA AGA AAC CTA GGA AAA TGT TGT AAA CCT (1430)							
441		448	450		460 461		470
pro glu ala lys arg met pro oys ala glu asp tyr leu ser val val leu gln leu cys val his glu lys thr arg val ser							
CCT GAA CCA AAA ACA ATC CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC CTC AAC CAC TTA TGT GTC CAT GAG AAA ACC CCA CTA ACT (1520)							
471	476 477	480		490		500	
asp arg val thr lys cys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val asp glu thre tyr val val oro lys							
GAC ACA GTC ACC AAA TGC TGC ACA GAA TCC TTC GTC AAC ACC CGA CCA CCT CCT GTC GAA GTC GAT GAA ACA TAC CCT CCC AAA (1610)							
501	510	514		520		530	
glu phe asn ala glu thr phe thr phs his ala asp lle cys thr leu ser glu lys gln lle lys lys qin thr ala leu val							
GAG TTT AAT GCT GAA ACA TTC ACC TTC CAT GCA GAT ATA TCC ACA CTT TCT GAG AAC ACA AAC ATC AAC ACT GCA CCT CTT CTT (1700)							

0079739

4083

-18-

5

10

15

20

25

30

35

531	540	550	560	
glu lys pro lys ala thr lys glu gln leu lys ala val met asp asp	ala ala phe val glu lys cys oys lys			
GAC GTC AAA CAC AAG CCC AAC ACA AAA GAC CAA CTG AAA GCT GTC TTT GAT CAT TCC GCT CCT TCT	AAC TCC TGC TAA TAA CATTAAAC			
561	567	570	580	
ala asp asp lys glu gln leu val ala ser gln ala leu gly leu ter				
CAT AAC CAC ACC TCC TTT CCC GAG GCT AAA AAA CTT GCT GCA ACT CAA GCT CCC TTA GCA TAA CATTAAAC				

CATCTAGCTTACCATGAGAAATTAGAGAAACAAAATGAAGTCAAAAACCTTATTCACTCGTTCTTCTTGTAAACCCAAACCTGTCTAAATTTCTTTAA (2002)

TCATTTGCCCTTCTCTGCTCAATTATAAGGAAAGCTA... 20 ...AA (20/8)

0079739

4083

-19-

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

5

10

15

20

25

30

35

GCTTTTCTCTCTCTGTAACCCACACCCCTTGGCACAAATG AAC TGG GCA ACC ATT TCC CTT ATT TCC CTT CTC TTT AGC (30)  
-18 p r o -10  
Met Lys trp val thr phe Ile ser leu leu phe Ile she Ser  
-1 -6 p r o -1  
ser ala tyr ser arg gly val phe arg arg  
TCG CCT TAT TCC AGG GGT GTC TTT CGT CCA

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

35	30	25	20	15	10	5
-6	P r p	-1	10	20	50	
arg gln val phe arg arg asp ala his lys ser glu val ala his arg phe lys asp leu gln glu asn phe lys						
AGC GGT GTC TTT CGT CCA CAT GCA CAC AAG AGT GAG CCT CAT CGG TTT AAA GAT TCC GCA GAA AAT TTC AAA (170)						
21	ala leu val ile ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val asn glu val thr glu phe ala					
GCC TTG CTC ATT CCC TTT GCT CAG TAT CTT CAC CAG TGT CCA AAA GAT CAT GTA AAA TTA GTC AAT GAA GCA ACT CAA TTT GCA (260)						
30	34	40	40	50		
51	53	60	62	70	75	80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly asp lys leu cys thr val ala thr leu						
AAA ACA TGT CTT CCT GAT GAC TCA GCT GAA AAT TGT GAC AAA TCA CTT CAT ACC CTT TTT CGA GAC AAA TTA TGC ACA GTC ACT CTT (350)						
90	91			100	101	
81	arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys aso asn pro					
CGT CAA ACC TAT GCT GAA ATC CCT GAC TGC ATG AGA CCA AAA CCT GCA TGT GCA AAA ACC TAT GAA TCC TTC TGC CAA CAC AAC GAC AAC CCA (660)						
120	124			130		140
111	asn leu pro arg leu val arg pro glu val met cys thr ala phe his aso asn glu glu thr ohe leu lys tyr leu try					
AAC CTC CCC CCA TTC GTC AGA CCA CAT CCT TAC TTT TAT GCC CGG GAA CTC CTT GTC TTT GCT AAA ACC TAT AAA CCT GCA TGT TCA TAT (330)						
150				160		168 169 170
141	glu ile ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ohe ala lys arg tyr lys ala ala phe thr glu cys gln					
CAA ATT CCC AGA CAT CCT TAC TTT TAT GCC CGG GAA CTC CTT GTC TTT GCT AAA ACC TAT AAA CCT GCA TGT TCA TAT (620)						
177	180			190		200
171	ala ala asp lys ala cys leu pro lys glu leu arg asp glu gly lys ala ser ser ala lys aln arg leu lys eys					
GCT GCT GAT AAA GCT GCC TGC CTC CCA AAG CTC GAT GAA CTT CGG GAT CAA GGG AGC GCT TCG TCT GCC AAA CAG AGA CTC AAC TGC (710)						
210				220		230
ala ser leu gln lys phe gly glu arg ala phe lys ala val ala arg leu ser gln arg ohe pro lys ala glu ohe ala glu						
CCC ACT CTC CAA AAA TTT GGA GAA AGA GCT TTC AAA GCA TCC GCA GAA GCT GCG CTC AGC CAG AGA TTT CCC AAA GCT GAC TTT GCA GAA (300)						

0079739

4083

-21-

35           38           25           20           15           10           "           5  
231        240        245 246        250        253        250        253        260  
val ser lys leu val thr asp leu thr lys val his thr glu cys eye his glu asp leu leu glu cys ala asp asp arg ala asp asp leu  
GTT TCC AAG TTA GTC ACA GAT CTT ACC AAA GTC AAC ACC GAA TGC CAT GCA GAT CTC CCT GAA TGT GCT GAT GAC AGC CGG GAC CTT (890)  
261        265        270        278 279 280        289 290  
ala lys tyr lle oys glu asn gln asp ser lle ser ser lle lys glu oys cys glu lys pro leu leu glu lys ser his cys lle  
GCC AAG TAT ATC TGT GAA AAT CAA GAT TCG ATC TCC ACT AAA CTC AAG GAA TGC TGT GAA AAA CCT CTG TTG GAA AAA TCC CAC TGC ATT (900)  
291        300        310        316        320  
ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys aso val cys lys asn tyr ala  
GCC GAA GTC GAA AAT GAT CAG ATG CCT CCT GAC TAC TGT CCT GCT GAT TTT GAT GAA AGT AAC TAC TAT GCT GTC AAA AAC TAT GCT (1070)  
321        330        340        350  
glu ala lys asp val phe leu gly met phe leu tyr ala arg arg his pro asp tyr ser val val leu leu ala  
GAG GCA AAG GAT GTC TTC TTG GCC ATG TTT TGC TAT GAA TAT GCA AGA ACC CAT CCT GAT TAC TCT GTC GTC CTC CTC CTC AGA CTT GCC (1160)  
351        360 361        369 370        380  
lys thr tyr glu thr leu glu lys cys cys ala ala ala asp pro his glu cys tyr ala lys val phe asp glu phe lys oru leu  
AAG ACA TAT GAA ACC ACT CTA GAG CCC GCT TGT GTC TGT GTC AAG TGC TCA ACT CCT GCA GAT CCT CAT GAA TGT TAT GCC AAA GTG TTG AAA CCT CCT (1250)  
381        390        392        400  
val glu glu oro gln asn leu lle lys gln ean cys glu leu phe alu gln leu gln ala leu leu gln asn ala leu leu val arg  
GTC GAA GAG CCT CAG AAA ATC TTA ATC AAA CAA AAT TGT GAG CAG CTT TTG GAC GAG CCT GCA GAG TAC AAA TTC CAG AAT CGG CTC TTA GTC CGT CGT (1360)  
411        420        430        440        450        460 461        470  
tyr thr lys lys val pro gln val ser thr pro thr leu val glu val ser arg asn leu ala lys val oly ser lys cys lys his  
TAC ACC AAG AAA GTA CCC CAA GTC TCA ACT CCT GCA GAG GTC TCA ACT CCA ACT CCT GCA GAT CCT GTC TGT AAA TGT TGT AAA TAT (1430)  
441        448        450        460 461        470  
pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val leu ean gln leu cys val leu his glu lys thr oru val ser  
CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA GAC TAT CTA TCC GTC GTC CTC AAC CAG TTA TGT GTC CAT GAC AAA AGG CCA GTC AGT (1520)  
471        476 477        480        490        500  
asp arg val thr lys cys thr glu ser leu val asn arg arg pro cys phe ser ala leu glu val aso olu thr tyr val pro lys  
GAC AGA GTC ACC AAA TCC TGC ACA GAA TCC TGC AAC AGC AAC ACC GCA CCA TGG TTT TCA CCT CTC GCA GTC CAT GAA ACA TAC GTC CCC AAA (1610)  
501        510        514        520        530  
glu phe asn ala glu thr phe thr phe his ala esp lle cys thr leu ser glu lys glu ero oln lle lys lys gln thr ala leu val  
GAG TTT AAT GGT GAA ACA TTC ACC TCC CAT GCA GAT ATA TGC ACA CTT TCT GAG AAC GAG AGA CAA ACT GCA CTT GTC (1700)

0079739  
4083

-22-

5

10

15

20

25

30

35

531 glu leu val lys his lys pro lys ala thr lys glu gln leu lys ala val met asp phe ala ala phe val glu lys cys cys lys  
GAG CTC GTC AAA CAC AAC CCC AAG GCA ACA AAA GAG CAA CTC AAA GCT GAT GAT ATG GAT GCT TTT GTC GAG AAG TGC TGC AAG (1790)  
540  
561 ala asp asp lys glu thr cys phe ala glu glu gln lys leu val ala ser gln ala ala leu gln leu ter  
GCT GAC GAT AAC GAG ACC TCC TTT GCC GAG CAG CTC AAA CTT GCT GCA AGT CAA GCT CCC TTA TAA CATCACATTAAAG (1883)  
570  
567 ter ter  
CATCTCAGCTTACCATGACAATAAGGAAAGAAAATGAGATCAAAGGETTATTCTCATCTGTTTCTTTCGTGTAAGGCCAACACCCCTGTCTAAACATAAATTCTTTAA (2078)  
TCATTTGGCTCTTCTCTGTGCTTCATTAAATAAAAATGGAAAGAICTAA.... 20 .....AA (2078)

0079739

4083

-23-

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

35	30	25	20	15	10	5
-1	-6	-1	-1	-10	-10	
Met lys trp val thr ile ser leu leu phe aer ATG AAC TGG GAA ACC TTT ATT TCC CTT CTC TTT AGC (30)						
21	16	11	10	20		
Ser ala tyr ser arg gly val phe ser arg asp s1a his lys ser glu val s1a his arg phe lys asp leu ala glu asn s1c lys TCC GCT TAT TCC ACC CCT GTC TTT CGT CCA GAT CCA CAC AAC AGT GAC GTC GCT CAT CGG TTT AAA GAT TTC GCA GAA GAA AAA (170)						
21	34	30	40	50		
ala leu val ile ilic ala phe ala gln tyr leu gln gln cys pro phe glu asp his val lys leu val ssn glu val thr glu phe ala GCC TTG GTC ATT GCC TTT CCT CAG TAT CTT CAC CAG TGT CCA TTT GAA CAT CAT GAA TTA GTC AAT CAA ACT CAA TTT GCA (1260)						
51	53	60	62	70	75	80
lys thr cys val ala asp glu ser ala glu asn cys asp lys ser leu his thr leu phe gly esp lys leu cys thr val ala thr leu AAA ACA TGT CTT CCT GAT CAG TCA GCT GAA AAA TGT GAC ACC CTT CTT CAT ACC CCT TTT CCA GAC AAA TTA TGC ACA GTC ACT CTT (350)						
81	90	91	100	101		
arg glu thr tyr gly glu met ala asp cys cys ala lys gln glu pro gly arg asn glu cys phe leu aln his lys asd asp asn pro CGT GAA ACC TAT GGT GAA ATG CCT GAC TCC TGT GCA AAA CAA CCT GCA AAT GAA CAC ACA TTT TGC TTC GAA TAC AAC CCA (440)						
111	120	124	130	140		
asn leu pro arg leu val arg pro glu val asp val met cys thr ala ala his asd asn glu glu thr phe leu lys tyr leu try AAC CTC CCC CCA TTC GTC ACA CCA CAG GTC GAT CCT GAT GTC ACT CCT TTT CAT GAC AAC TAT GAA GAC ACA TTT TGC CAA (330)						
141	150	154	160	168 169 170		
glu ilic ala arg arg his pro tyr phe tyr ala pro glu leu leu phe ala lys arg tyr lys ala ala phe thr glu cys qln GAA ATT GGC ACA ACA CAT CCT TAC TTT TAT GGC CCG GAA CTC CTT TTC TTT GCT AAA AGG TAT AAA GCT CCT TTT ACA GAA TGT TGC CAA (620)						
171	177	180	190	200		
ala ala asp lys ala ala cys leu pro lys leu esp glu leu arg esp glu ala lys ala ser ser ala lys aln arg leu lys cys CCT GCT GAT AAA CCT GCT CCC TCC CTG TTG CCA AAG CTC GAT GAA CTT CGG GAT GAA GCG AAC GCT TCC TCT CCC AAA CAC ACA CTC AAC TGT (710)						
201	210	220	230			
ala ser leu gln lys phe gly glu arg ala phe lys ala trp ala val ala arg leu ser gln arg phe pro lys ala glu phe ala glu GCC ACT CTC CAA AAA TTT GCA GAA GCA GCT TTT CCC AAA CCT CCT CTC ACC CAG AGA TGA TTT GCA GAA (3000)						

0079739

4083

-24-

231            33            240            25            246            265            246            250            253  
 val ser lys leu val thr asp leu thr lys val his thr glu oys oys his gly asp leu glu cys ala asp arg ala asp leu  
 GTC TCC AAC TTA GTC ACA CAT CCT ACC AAA GTC CAC ACC GAA TCC TGC CAT GCA GAT CTC CTT GAA TGT GCT CAT GAC AGC GCG GAC CTT (1890)  
  
 261            265            270            278            279 280            289 290  
 ala lys tyr lle cys glu asn gln asp ser lle ser ser lys leu lys glu cys cys glu lys pro leu glu lys ser.his cys lle  
 CCC AAC TAT ATC TGT GAA AAT CAA GAT TCC ATC TCC ACT AAA CTC AAC TGC TGT GAA AAA CCT CTC TGC TGT GAA AAC TAT GCT ATT (1980)  
  
 291            300            310            316            320  
 ala glu val glu asn asp glu met pro ala asp leu pro ser leu ala ala asp phe val glu ser lys asp val cys lys asn tyr ala  
 GCC GAA CTG GAA AAT GAT GAC ATG CCT GCT GCT GAC TTG CCT TCA TTA GCT GAT TTT GTC GAA AGT AAG GAT GCT ATT GTC AAA AAC TAT GCT ATT (11070)  
  
 321            330            340            350  
 glu ala lys asp val phe leu gly met phe leu tyr ala arg arg his pro asp tyr ser val val lys ala asp tyrosine  
 GAC GCA AAC GAT GTC TTG CCC ATG TTT TTG GGC AAC TAC CCT GAA TAT GCA AGC CAT CCT GAT TAC TCT GTC GTC CTC CTC GAA CCC (11601)  
  
 351            360 361            369 370            380  
 lys thr tyr glu thr leu glu lys cys ala ala ala asp pro his glu cys tyr ala lys val one asp glu cys tyr ala lys  
 AGG ACA TAT GAA ACC ACT CTA GAC AAC TGC TGT GCT GCA GAT CCT CAT GAA TGC TAT GCC AAA GTC TTC GAT GAA TTT AAA CCT CCT CGT (13250)  
  
 381            390            392            400  
 val glu glu pro gln asn leu lle lys gln asn cys glu leu phe glu gln leu gly glu tyr lys phe gln asn ala leu leu val  
 GTC GAA GAG CCT CAG AAT TTA ATC AAA CAA ACT CCT GIA GAG GTC TCA ACT CCA ACT CCT GIA GAG GTC GCT TCA AGA AAC CTA GGA AAA GTG GAC AGC (14300)  
  
 411            420            430            440  
 tyr thr lys lys val pro gln val val lys val gln leu gln lys val gln leu gln lys ser arg asn leu gln lys  
 TAC ACC AAG AAA GTC CCC CAA GTG TCA ACT CCT GIA GAG GTC TCA ACT CCA ACT CCT GIA GAG GTC GCT TCA GCT TGT AAA CAT (15300)  
  
 461            468            450            470            471  
 pro glu ala lys arg met pro cys ala glu asp tyr leu ser val val lys val gln leu cys val leu glu val asp glu thr  
 CCT GAA GCA AAA AGA ATG CCC TGT GCA GAA CAC TAT CTA TCC GTC AAC CAG TTA TGT GTC TGT GCA GAA TCC GTC GAA ACA TAC GCA GCT (16100)  
  
 476 477            480            490            500  
 asp arg val thr lys oys cys thr glu ser leu val asn gln leu cys val leu glu val asp glu thr tyr val pro lys  
 GAC AGA GTC ACC AAA TCC TGC ACA GAA TCC TTC GTC GTC AAC AGC CAG AGA CAA ATC AAC AAA GCA ACT GCA CTT GTC GAA (17000)  
  
 510            514            520            530  
 glu phe asn ala glu thr phe thr phe his ala asp lle cys thr leu ser glu lys glu arg aln lle lys lys oln thr ala leu val  
 GAC TTT ATT GCT GAA ACA ATT TCC ACA CCT TCT GTC AAC GAG AGA CAA ATC AAC AAA GCA ACT GCA CTT GTC GAA (5300)

0079739

4083

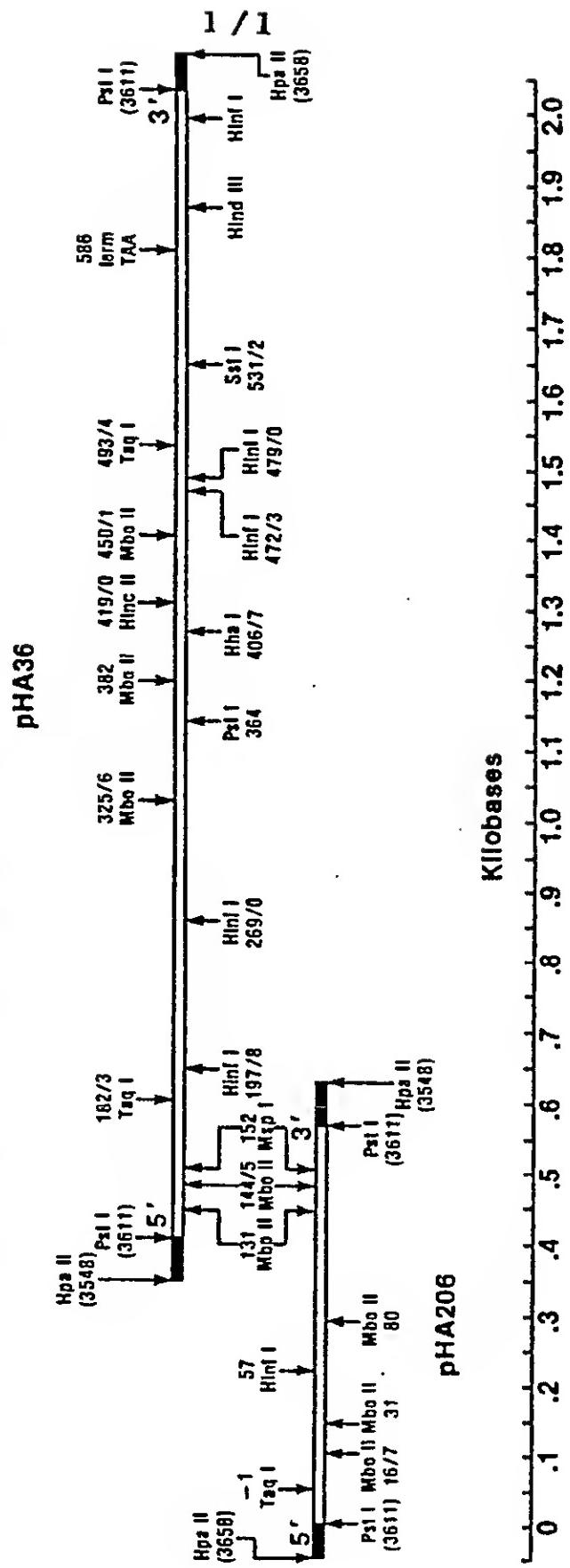
-25-

5  
10  
15  
20  
25  
30  
35

10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
13. A DNA transfer vector according to claim 12, which is a plasmid.
14. A DNA transfer vector according to claim 13,
- 10 wherein the plasmid is pBR322 or YEp6.
15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
17. A vector or process according to claim 16, wherein the bacterium or yeast is E. coli or Saccharomyces cerevisiae.

0079739

Restriction Endonuclease Map of Human Serum Albumin cDNA Clones



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER: \_\_\_\_\_**

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.